

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2002, 00:09:17 ; Search time 8498.8 Seconds
(without alignments)
31.610 Million cell updates/sec

Title: US-09-851-670-6
Perfect score: 25
Sequence: 1 cccctagcccccacacgactactactgct 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 111874

Minimum DB seq length: 0
Maximum DB seq length: 60

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estcpl:*
6: em_estlba:*
7: em_estlro:*
8: em_estlov:*
9: em_hic:*
10: gb_estl:*
11: gb_estl2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	63.2	54	13	A2766014 IM0563H17
2	15.6	62.4	52	13	A2327082 IM0050D15
3	14.6	58.4	43	13	A2762861 IM0558P04
4	14.6	58.4	57	10	AA291683 zt39d03.s
5	14.2	56.8	43	13	A1048043 vn21f12.x
6	13.8	55.2	45	13	A2599477 IM0414B20
7	13.8	55.2	50	10	AU104177 AU104177
8	13.8	55.2	50	10	AU104182 AU104182
9	13.8	55.2	50	10	AU104190 AU104190
10	13.6	54.4	27	13	A2774487 2M0004D01
11	13.6	54.4	56	11	BF018760 ux97h05.x
12	13.4	53.6	42	13	A2370282 IM0121C16

13	13.2	52.8	28	13	A2586097
14	13.2	52.8	43	11	BE966209
15	13	52.0	32	13	A2581120
16	12.8	51.2	40	10	AA811470
17	12.8	51.2	50	10	AU102901
18	12.8	51.2	50	10	AU102902
19	12.8	51.2	50	10	AU105107
20	12.8	51.2	50	10	AU105124
21	12.8	51.2	55	10	AU105170
22	12.6	50.4	46	10	AA867748
23	12.6	50.4	50	10	AU105885
24	12.6	50.4	52	11	H14777
25	12.4	49.6	34	13	A2586611
26	12.4	49.6	50	10	AU102899
27	12.4	49.6	50	10	AU102900
28	12.4	49.6	50	10	AU107387
29	12.4	49.6	50	10	AU107388
30	12.4	49.6	50	10	AU107389
31	12.4	49.6	59	10	AA419276
32	12.4	49.6	60	10	A1984456
33	12.2	48.8	21	13	A2596843
34	12.2	48.8	41	13	A2775318
35	12.2	48.8	46	10	AA985334
36	12.2	48.8	49	10	AA011834
37	12.2	48.8	50	10	AU104186
38	12.2	48.8	51	13	A2591236
39	12.2	48.8	52	11	H39343
40	12.2	48.8	54	13	A2576640
41	12.2	48.8	55	10	AA894695
42	12.2	48.8	56	13	A2775455
43	12.2	48.8	58	10	A1095308
44	12.2	48.8	58	10	A1375271
45	12.2	48.8	60	13	A2456310

ALIGNMENTS

RESULT 1
LOCUS A2766014 54 bp DNA
DEFINITION IM0563H17F Mouse 10kb plasmid UGCCIM library Mus musculus genomic
clone UGCCIM0563H17 F, DNA sequence.
ACCESSION A2766014
VERSION A2766014.1 GI:12882637

KEYWORDS
GSS.
SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
1 (bases 1 to 54)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0563 row: H column: 17
Seq primer: CGTGTAAACGACGCCGCT
Class: Plasmid ends
High quality sequence stop: 54.
Location/Qualifiers
1. 54

FEATURES
source

```

BASE COUNT
ORIGIN
11 a
16 c
13 g
14 t

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0563H17"
/clone_1fb="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42mv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g1147321141gb1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

Query Match	63.2%	Score 15.8	DB 13	Length 54
Best Local Similarity	89.5%	Pred. No. 1,2e+04		
Matches 17; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	ccttagcggccacacagctct	19	
DB	18	cccaaggccccaacacagctct	36	

RESULT	2
AZ327082	
LOCUS	AZ327082 52 bp DNA
DEFINITION	I06050D15 Mouse 10kb plasmid UUGC1M library GSS clone UUGC1M050D15 F, DNA sequence.
ACCESSION	Az327082
VERSION	AZ327082.1 GI:10385476
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Mus; 1 (bases 1 to 52)
AUTHORS	Dunn, P., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinger, A., von Niederhausen, A. and Wright, D., Weiss, R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL	Plasmid Inserts
COMMENT	Unpublished (2000) Contact: Robert B. Weiss

University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCC, UT
 84112, USA
 Tel.: 801 585 5606
 Fax: 801 585 7177
 Email: ddunne@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0050 row: D column: 15
 Seq primer: CGTTGTAAMACGACGCCACAGT
 Class: plasmid ends
 High quality sequence stop: 52.
 Location/Qualifiers

```

source
1. .52
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UNGClM0050D15"
/clone_lib="Mouse 10kb plasmid UNGClM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g147321141gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
16 a 13 c 11 g 12 t

```

Query Match	62.4%	Score 15.6	DB 13	Length 52
Best Local Similarity	81.8%	Pred. No. 1.5e+04		
Matches	18	Conservative	0	Mismatches 4
				Indels 0
				Gaps 0
QY	1	ccttaggcctccaccagcttact	22	
DB	11	cccaagaccacacccagctcaact	32	

RESULT	3
AZ762861	
LOCUS	43 bp DNA
DEFINITION	M0558P04.F Mouse 10 kb plasmid UNGC1M library Mus musculus genomic
ACCESSION	clone UNGC1M0558P04 F, DNA sequence.
VERSION	AZ762861.1 GI:12873302

ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
Mus musculus	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hambl, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts	Unpublished (2000)	Contact: Robert B. Weiss
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	1 (bases 1 to 43)			

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0558 row: P column: 04
Seq primer: CGTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 43.


```

/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pSPORT; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dT).
5'-CGGTGACCGTCGACCGCTTTTCTTTTCTTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
BASE COUNT      13 a      5 c      16 g      9 t
ORIGIN
Query Match      56.8%; Score 14.2; DB 10; Length 43;
Best Local Similarity 84.2%; Pred. No. 5.2e+04;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy      5 agagccacacgcgtctacag 23
          ||||||| ||||| |||||
Db      42 agcccccctccagcgtactg 24

```

RESULT	6
AZ599477/c	
LOCUS	AZ599477 45 bp DNA GSS 13-DEC-2000
DEFINITION	M0414B20 Mouse 10kb plasmid UGCI1M library Mus musculus genomic clone UGCIM0414B20 R. DNA sequence.
ACCESSION	AZ599477
VERSION	AZ599477.1 GI:1172167
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 45)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

TITLE	Mouse whole genome scaffolding with paired end reads from 10Kb
JOURNAL	plasmid Insetis
COMMENT	Unpublished (2000)
	Contact: Robert B. Weiss

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0414 row: B column: 20
Seq primer: CACACAGCAACACACTGTGACC
Class: plasmid ends
High quality sequence stop: 45.

FEATURES
source

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGGCM0414820"
/clone_1kb="Mouse 10kb plasmid UGGCM library"
/sex="Male"
/lab_host="E. Coli strain XL10-GOLD, T1-resistant, F-"
/notes="Vector: PWDΔuvr; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrothermally sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The

```

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1147321141gb1AR129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match	Score 13.8;	DB 13;	length 45;
Best Local Similarity	72.0%;	Pred. No. 7.5e+04;	
Matches 18;	Conservative 0;	Mismatches 7;	Indels 0;
		Gaps 0;	
Qy	1 ccctagagcccccacagctactgct	25	
Db	29 cccatagcccccacccgitttcttct	5	

[illegible]

REFERENCE
AUTHORS
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominae: Homo.
1 (bases 1 to 50)
Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata

FEATURES	JOURNAL COMMENT	TITLE
<p>Location/Qualifiers</p> <p>5' end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).</p>	<p>Contact: Yutaka Suzuki Department of Virology Institute of Medical Science, University of Tokyo 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan Email: yanzuk@iims.u-tokyo.ac.jp</p>	<p>Fine Structural analysis of transcription start sites of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries Unpublished (2001)</p>

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HS104173"
/clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT      11 a      20 c      6 g      13 t
ORIGIN

```

Query Match	55.2%	Score 13.8	DB 10	length 50
Best Local Similarity	72.0%	Pred. No. 7.5e+04		
Matches 18: Conservative	0: Mismatches	7: Indels	0: Gaps	0:
Oy	1	ccctaagccaccagctactactgct	25	
db	6	ccctggcccaaccgctcactact	30	

RESULT	8
LOCUS	AU104182
DEFINITION	AU104182 50 bp mRNA EST 05-APR-2001 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone

ACCESSION KAT02326, mRNA sequence.
 AU104182
 VERSION AU104182.1 GI:13553703
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 50)

REFERENCE Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
 H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
 K., Suyama,A. and Sugano,S.
 TITLE Fine structural analysis of transcription start sites of human
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries
 JOURNAL unpublished (2001)
 COMMENT Contact: Yutaka Suzuki
 Department of Virology
 Institute of Medical Science, University of Tokyo
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
 Email: yusuzuki@ms.u-tokyo.ac.jp

FEATURES
 source 1..50
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="KAT02326"
 /clone_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 11 a 20 c 6 g 13 t
 ORIGIN

Query Match 55.2%; Score 13.8; DB 10; Length 50;
 Best Local Similarity 72.0%; Pred. No. 7.5e+04;
 Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 cccatggccccacagctactgct 25
 DB 7 CCCGTGGCCCAACCGTCATCTACT 31

RESULT 9
 LOCUS AU104190 50 bp mRNA EST 05-APR-2001
 DEFINITION AU104190 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
 LNC15187, mRNA sequence.
 ACCESSION AU104190
 VERSION AU104190.1 GI:13553711
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 50)

REFERENCE Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
 H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
 K., Suyama,A. and Sugano,S.
 TITLE Fine structural analysis of transcription start sites of human
 mRNAs using full-length enriched and 5'-end enriched cDNA libraries
 JOURNAL unpublished (2001)
 COMMENT Contact: Yutaka Suzuki
 Department of Virology
 Institute of Medical Science, University of Tokyo
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
 Email: yusuzuki@ms.u-tokyo.ac.jp

FEATURES
 source 1..50
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="KAT02326"
 /clone_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 16 a 17 c 6 g 11 t
 ORIGIN

Query Match 55.2%; Score 13.8; DB 10; Length 50;
 Best Local Similarity 72.0%; Pred. No. 7.5e+04;
 Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 cccatggccccacagctactgct 25
 DB 22 CCCGTGGCCCAACCGTCATCTACT 46

RESULT 10
 LOCUS A2774487 27 bp DNA GSS 16-FEB-2001
 DEFINITION 2M0004D01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC2M0004D01 F, DNA sequence.
 ACCESSION A2774487
 VERSION A2774487.1 GI:12899988
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 27)

REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
 and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert Length: 1000 Std Error: 0.00
 Plate: 0004 row: D column: 01
 Seq primer: CGTTGTAACGACGCGCAGT
 Class: plasmid ends
 High quality sequence stop: 27.
 Location/Qualifiers

FEATURES
 source 1..27

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0004D01"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/nanres/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g14732114|9b|aE129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to

BASE COUNT 5 a 2 c 12 g 8 t
 ORIGIN adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match 54.4%; Score 13.6; DB 13; Length 27;
 Best Local Similarity 80.0%; Pred. No. 8.8e+04;
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 5 agggccaccagctactgc 24
 |||||||||
 Db 20 AGACCCACCTCAATTGC 1

RESULT 11
 BF018760 56 bp mRNA EST 29-DEC-2000
 LOCUS ux97h05.x1 Mccarrey Eddy spermatoocytes Mus musculus cDNA clone
 DEFINITION IMAGE:3656505 3', mRNA sequence.
 ACCESSION BF018760 GI:10750092
 VERSION BF018760
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 56)
 Author(s): Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Rheising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Korn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Other ESTs: ux97h05.y1
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:1418809
 COMMENT Seq primer: Primer name ambiguous.

FEATURES

source Location/Qualifiers
 1..56
 /organism="Mus musculus"
 /strain="CD-1"
 /db_xref="taxon:10090"
 /clone="IMAGE:3656505"
 /clone_1lb="Mccarrey Eddy spermatoocytes"
 /sex="male"
 /tissue_type="spermatoocytes, pooled from multiple mice"
 /dev_stage="60 day"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: testis; Vector: pBluescript SK+ (Stratagene); Site 1: XhoI; Site 2: EcoRI; cDNA oligo dt-primed 15'-(GA)10-ACGAGTCGAGTTTCTTTT-3' and directionally cloned using 5' linkers 5'-AATTGCGCAGAG-3' and 5'-CTCGTGGCG-3'. Size selection of >400bp material gives average insert size ranging from 1-2 kb. Library was mass excised (from lambda-Unizap-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library contains 98% recombinants.
 References: J. Androl. 20:635-639 and Gene 25:263-269.
 Library constructed and donated by J. Mccarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is

BASE COUNT 15 a 13 c 11 g 17 t
 ORIGIN available through ATCC, catalog #63422."

Query Match 54.4%; Score 13.6; DB 11; Length 56;
 Best Local Similarity 80.0%; Pred. No. 9.1e+04;
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 ccctagggccaccagctca 20
 |||||||||
 Db 33 CACTGGGCCCCACGAGCAA 52

RESULT 12
 A2370282 42 bp DNA GSS 02-OCT-2000
 LOCUS 1M0121C16F Mouse 10kb plasmid UUGC1M 1library Mus musculus genomic clone UUGC1M0121C16 F, DNA sequence.
 DEFINITION A2370282
 ACCESSION A2370282
 VERSION A2370282
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 42)
 Author(s): Dunn, D., Aoyagi, A., Barber, M., Becorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A., and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0121 row: C column: 16
 Seq primer: CGTTGTAAACGACGCGCAGT
 Class: plasmid ends
 High quality sequence stop: 42.

FEATURES

source Location/Qualifiers
 1..42
 /organism="Mus musculus"
 /strain="C57BL/6j"
 /db_xref="taxon:10090"
 /clone="UUGC1M0121C16"
 /clone_1lb="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6j (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114(g1b)AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into

chemically-competent *E. coli* XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 8 a 4 c 14 g 16 t

ORIGIN

Query Match 53.6%; Score 13.4; DB 13; Length 42;
Best Local Similarity 73.9%; Pred. No. 1.1e+05;
Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 ctaggcccccagctctact 25
|||||
Db 34 CTAGCCTCCAGCATTAACGTCT 12

RESULT 13

AZ586097 28 bp DNA GSS 13-DEC-2000
LOCUS 1M0391G17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0391G17 R, DNA sequence.
ACCESSION AZ586097
VERSION AZ586097.1 GI:11708287
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
COMMENT Unpublished (2000)
CONTACT: Robert B. Weiss
UNIVERSITY of Utah Genome Center
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0391 row: G column: 17
Seq primer: CACACAGGAAACAGCTATCACC
Class: plasmid ends
High quality sequence stop: 28.

FEATURES

Source
1.28
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0391G17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g11473211419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to

adaptor vector DNA, and transformed into
chemically-competent *E. coli* XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 6 a 11 c 3 g 8 t

ORIGIN

Query Match 52.8%; Score 13.2; DB 13; Length 28;
Best Local Similarity 83.3%; Pred. No. 1.3e+05;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cctaggcccccagctct 19
|||||
Db 11 CCTGTGCCCCACCATCTCT 28

RESULT 14

BE896209 43 bp mRNA EST 20-OCT-2000
LOCUS 601438919F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3923716 5',
DEFINITION mRNA sequence.
ACCESSION BE896209
VERSION BE896209.1 GI:10360382
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 43)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9760 row: f column: 05
High quality sequence stop: 43.

FEATURES

Source
1.43
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3923716"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 11 a 11 c 11 g 10 t

ORIGIN

Query Match 52.8%; Score 13.2; DB 11; Length 43;
Best Local Similarity 83.3%; Pred. No. 1.3e+05;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 agggcccccagctctact 22
|||||
Db 10 AGCTCCACACGCGCTATT 27

RESULT 15

AZ581120 32 bp DNA GSS 13-DEC-2000
LOCUS 1M0369E02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0369E02 R, DNA sequence.

ACCESSION AZ581120
 VERSION AZ581120.1 GI:11695814

KEYWORDS GSS.

SOURCE

ORGANISM

house mouse.

Mus musculus

REFERENCE

1 (bases 1 to 32)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A., and Wright, D., Weis, R.

AUTHORS

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL

unpublished (2000)

COMMENT

Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177

Email: ddunne@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0369 row: E column: 02

Seq primer: CACACAGGAACAGCTATGACC

Class: Plasmid ends

High quality sequence stop: 32.

FEATURES

source

Location/Qualifiers

1..32

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUCG1M0369E02"

/clone_lib="Mouse 10kb plasmid UUCG1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PMD42 (g11473211419b1AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 8 a 9 c 9 g 6 t

ORIGIN

Query Match 52.0%; Score 13; DB 13; Length 32;
 Best Local Similarity 76.2%; Pred. No. 1.5e+05;
 Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 agggccaccagctctactgct 25
 ||||| ||||| ||||| |||||
 DB 3 AGGCACTACCTGACACTGCT 23